Date: June 8, 2000 Express Mail Label No. EL564265605US

Inventors: Guy A. Rouleau and Bernard Brais

Attorney's Docket No.: 3028.1000-000

SHORT GCG EXPANSIONS IN THE PAB II GENE FOR OCULO-PHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

RELATED APPLICATION(S)

This application is a continuation claiming priority to International Application No. PCT/CA98/01133, filed December 7, 1998 (designating the U.S.), which claims priority to Canadian Patent No. 2,218,199, filed December 9, 1997, the entire teachings of both of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to PAB II gene, and its uses thereof for the diagnosis, prognosis and treatment of a disease related with protein accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

Description of Prior Art

Autosomal dominant oculopharyngeal muscular dystrophy(OPMD) is an adult-onset disease with a world-wide distribution. It usually presents itself in the sixth decade with progressive swallowing difficulties (dysphagia), eye lid drooping (ptosis) and proximal limb weakness. Unique nuclear filament inclusions in skeletal muscle fibers are its pathological hallmark (Tome, F.M.S. & Fardeau, Acta Neuropath. 49, 85-87 (1980)). Using the full power of linkage analysis in eleven French Canadian families,

the oculopharyngeal muscular dystrophy gene was fine mapped on human chromosome 14 (Brais et al., 1997, Neuromuscular Disorders 7 (Suppl.1):S70-74). A region of .75 cM was thereby identified as a region containing the potential and unknown OMPD gene (Brais et al., 1997, *supra*). Unfortunately, the OMPD gene has yet to be isolated and its nucleic acid or protein sequence have yet to be cribbed.

It would be highly desirable to be provided with a tool for the diagnosis, prognosis and treatment of a disease related with polyalanine accumulation in the nucleus, such as observed in oculopharyngeal muscular dystrophy.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide a tool for the diagnosis, prognosis and treatment of a disease related with polyalanine accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

Herein, the poly(A) binding protein II (PAB II) gene was isolated from a 217 kb candidate interval in chromosome 14q11.A (GCG)6 repeat encoding a polyalanine tract located at the N-terminus of the protein was expanded to (GCG)8-13 in the 144 OPMD families screened. More severe pheno-types were observed in compound heterozygotes for the (GCG)9 mutation and a (GCG)7 allele found in 2% of the population, whereas homozygosity for the (GCG)7 allele leads to autosomal recessive OPMD. Thus the (GCG)7 allele is an example of a polymorphism which can act as either a modifier of a dominant phenotype or as a recessive mutation. Pathological expansions of the polyalanine tract may cause mutated PAB II oligomers to accumulate as filament inclusions in nuclei.

In accordance with the present invention there is provided a human PAB II gene containing a transcribed polymorphic GCG repeat, which comprises a sequence as set forth in Fig. 4, which includes introns and flanking genomic sequence.

The allelic variants of GCG repeat of the human PAB II gene are associated with a disease related with protein accumulation in the nucleus, such as polyalanine

accumulation, or with a disease related with swallowing difficulties, such as oculopharyngeal muscular dystrophy.

In accordance with the present invention there is also provided a method for the diagnosis of a dis-ease associated with protein accumulation in the nucleus, which comprises the steps of:

a) obtaining a nucleic acid sample of said patient; and

b) determining allelic variants of a GCG repeat of the human PAB II gene; thereby long allelic variants are indicative of a disease related with protein accumulation in the nucleus, such as polyalanine accumulation and oculopharyngeal muscular dystrophy.

The long allelic variants have from about 245 to about 263 bp in length.

In accordance with the present invention there is also provided a non-human mammal model for the human PAB II gene, whose germ cells and somatic cells are modified to express at least one allelic variant of the PAB II gene and wherein said allelic variant of the PAB II is being introduced into the mammal, or an ancestor of the mammal, at an embryonic stage.

In accordance with the present invention there is also provided a method for the screening of thera-peutic agents for the prevention and/or treatment of oculopharyngeal muscular dystrophy, which comprises the steps of:

a) administering the therapeutic agents to the non-human animal of the present invention or oculopharyngeal muscular dystrophy patients; and

b) evaluating the prevention and/or treatment of development of oculopharyngeal muscular dystro-phy in this animal (such as a mammal) or in patients.

In accordance with the present invention there is also provided a method to identify genes-products thereof, or part thereof, which interact with a biochemical pathway affected by the PAB II gene, which comprises the steps of:

a) designing probes and/or primers using the PAB II gene and screening oculopharyngeal muscular dystrophy patients samples with said probes and/or primers; and

b) evaluating the role of the identified gene in oculopharyngeal muscular dystrophy patients.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1B illustrate the positional cloning of the PAB II gene;

Sub Q1 Figs. 2A-2G illustrate the OPMD (GCG)n expansion sizes and sequence of the mutation site (SEQ ID NOS:1-2);

Fig. 3 illustrates the age distribution of swallowing time (st) for French Canadian OPMD carriers of the (GCG)9 mutation; and

Sub Q2 Figs. 4A-4E illustrate the nucleotide sequence of human poly(A) binding protein II (hPAB II)(SEQ ID NO:3).

DETAILED DESCRIPTION OF THE INVENTION

In order to identify the gene mutated in OPMD, a 350 kb cosmid contig was constructed between flanking markers D14S990 and D14S1457 (Fig. 1A). Positions of the PAB II-selected cDNA clones were determined in relation to the EcoRI restriction map and the Genealogy-based Estimate of Historical Meiosis (GEHM)-derived candidate interval (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of tran-scribed sequences (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

The human poly(A) binding protein II gene (PAB II) is encoded by the nucleotide sequence as set forth in Fig. 4.

Sub 68 Twenty-five cDNAs were isolated by cDNA selection from the candidate interval (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of transcribed sequences (eds. Hochgeschwender, U. & Gardiner, K.; GW 2/27/02

65-79; Plenum, New York, 1994). Three of these hybridized to a common 20 kb EcoRI restriction fragment and showed high sequence homology to the bovine poly(A) binding protein II gene(bPAB II) (Fig. 1A). The PAB II gene appeared to be a good candidate for OPMD because it mapped to the genetically defined 0.26 cM candidate interval in 14q11 (Fig. 1A), its mRNA showed a high level of expression in skeletal muscle, and the PAB II protein is exclu-sively localized to the nucleus (Krause, S. et al., Exp. Cell Res. 214, 75-82 (1994)) where it acts as a factor in mRNA polyadenylation (Whale, E., Cell 66, 759-768 (1991); Whale, E. et al., J. Biol. Chem. 268, 2937-2945 (1993); Bienroth, S. et al., EMBO J. 12, 585-594 (1993)).

A 8 kb HindIII genomic fragment containing the PAB II gene was subcloned and sequenced (6002 bp; GenBank: AF026029)(Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)) (Fig. 1B). Genomic structure of the PAB II gene, and position of the OPMD (GCG)n expansions. Exons are numbered. Introns 1 and 6 are variably present in 60% of cDNA clones. ORF, open reading frame; cen, centromere and tel, telomere.

The coding sequence was based on the previously published bovine sequence (GenBank: X89969) and the sequence of 31 human cDNAs and ESTs. The gene is composed of 7 exons and is transcribed in the cen-ter orientation (Fig. 1B). Multiple splice variants are found in ESTs and on Northern blots (Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). In particular, introns 1 and 6 are present in more than 60% of clones (Fig. 1B)(Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). The coding and protein sequences are highly conserved between human, bovine and mouse (GenBank: U93050). 93% of the PAB II sequence was readily amenable to RT-PCR- or genomic-SSCP screening. No mutations were uncovered using both techniques. However, a 400 bp region of exon 1 containing the start codon could not be readily amplified. This region is 80% GC rich. It includes a (GCG)6 repeat which codes for the first six alanines of a homopolymeric stretch of 10 (Fig. 2G). Nucleotide sequence of the mutated region of PAB II as well as the amino acid

sub B10

sequences of the N-terminus polyalanine stretch and position of the OPMD alanine insertions is also shown in Fig. 2.

Special conditions were designed to amplify by PCR a 242 bp genomic fragment including this GCG-repeat. The (GCG)6 allele was found in 98% of French Canadian non-OPMD control chromosomes, whereas 2% of chromosomes carried a (GCG)7 polymorphism (n=86) (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)).

Screening OPMD cases belonging to 144 families showed in all cases a PCR product larger by 6 to 21 bp than that found in controls (Fig. 2A). (GCG)6 normal allele (N) and the six different (GCG)n expansions observed in 144 families.

sub B11

Sequencing of these fragments revealed that the increased sizes were due to expansions of the GCG repeat (Fig. 2G). Fig. 2F shows the sequence of the (GCG)9 French Canadian expansion in a heterozygous par-ent and his homozygous child. Partial sequence of exon 1 in a normal (GCG)6 control (N), a heterozygote (ht.) and a homozygote (hm.) for the (GCG)9-repeat mutation. The number of families sharing the different (GCG)n-repeats expansions is shown in Table 1.

Table 1

Number of families sharing the different dominant (GCG)n OPMD mutations

Mutations	Polyalanine†	Families
(GCG)8	12	4
(GCG)9	13	99
(GCG)10	14	19
(GCG)11	15	16
(GCG)12	16	5
(GCG)13	17	1
Total		**144**

†, 10 alanine residues in normal PAB II.

The (GCG)9 expansion shared by 70 French Cana-dian families is the most frequent mutation we observed (Table 1).The (GCG)9 expansion is quite stable, with a single doubling observed in family F151 in an estimated 598 French Canadian meioses (Fig. 2C). The doubling of the French Canadian (GCG)9 expansion is demonstrated in Family F151.

This contrasts with the unstable nature of previously described disease-causing triplet-repeats (Rosenberg, R.N., New Eng. J. Med. 335, 1222-1224 (1996)).

Genotyping of all the participants in the clinical study of French Canadian OPMD provided molecular insights into the clinical variability observed in this condition. The genotypes for both copies of the PAB II mutated region were added to an anonymous version of this clinical database of 176 (GCG)9 mutation carriers (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)). Severity of the phenotype can be assessed by the swallowing time (st) in seconds taken to drink 80 cc of ice-cold water (Brais, B.

et al., Hum. Mol. Genet. 4, 429-434 (1995); Bouchard, J.-P. et al., Can. J. Neurol. Sci. 19, 296-297 (1992)).The late onset and progressive nature of the muscular dystrophy is clearly illustrated in heterozygous carriers of the (GCG)9 mutation (bold curve in Fig. 3) when compared to the average st of control (GCG)6 homozygous participants(n=76, thinner line in Fig. 3). The bold curve represents the average OPMD st for carriers of only one copy of the (GCG)9 mutation (n=169), while the thinner line corresponds to the average st for (GCG)6 homozygous normal controls(n=76). The black dot corresponds to the st value for individual VIII. Roman numerals refer to individual cases shown in Figs. 2B, 2D and discussed in the text. The genotype of a homozygous (GCG)9 patient and her parents is shown in Fig. 2B. Independent segregation of the (GCG)7 allele is also shown. Of note, case V has a more severe OPMD phenotype (Fig. 2D).

Two groups of genotypically distinct OPMD cases have more severe swallowing difficulties. Individuals I, II, and III have an early-onset disease and are homozygous for the (GCG)9 expansion ($P < 10\text{-}5$) (Figs. 2B, F). Cases IV, V, VI and VII have more severe phenotypes and are compound heterozygotes for the (GCG)9 mutation and the (GCG)7 polymorphism ($P < 10\text{-}5$). In Fig. 2D the independent segregation of the two alleles is shown. Case V, who inherited the French Canadian (GCG)9 mutation and the (GCG)7 polymorphism, is more symptomatic than his brother VIII who carries the (GCG)9 mutation and a normal (GCG)6 allele (Figs. 2D and 3). The (GCG)7 polymorphism thus appears to be a modifier of severity of dominant OPMD. Furthermore, the (GCG)7 allele can act as a recessive mutation. This was documented in the French patient IX who inherited two copies of the (GCG)7 polymorphism and has a late-onset autosomal recessive form of OPMD (Fig. 2E). Case IX, who has a recessive form of OPMD, is shown to have inherited two copies of the (GCG)7 polymorphism.

This is the first description of short trinu-cleotide repeat expansions causing a human disease. The addition of only two GCG repeats is sufficient to cause dominant

OPMD. OPMD expansions do not share the cardi-nal features of "dynamic mutations". The GCG expansions are not only short they are also meiotically quite sta-ble. Furthermore, there is a clear cut-off between the normal and abnormal alleles, a single GCG expansion causing a recessive phenotype. The PAB II (GCG)7 allele is the first example of a relatively frequent allele which can act as either a modifier of a dominant pheno-type or as a recessive mutation. This dosage effect is reminiscent of the one observed in a homozygote for two dominant synpolydactyly mutations. In this case, the patient had more severe deformities because she inher-ited two duplications causing an expansion in the polyalanine tract of the HOXD13 protein (Akarsu, A.N. et al., Hum. Mol. Genet. 5, 945-952 (1996)). A duplication causing a similar polyalanine expansion in the a subunit 1 gene of the core-binding transcription factor (CBF(1) has also been found to cause dominant cleido-cranial dysplasia (Mundlos, S. et al., Cell 89, 773-779 (1997)). The mutations in these two rare diseases are not triplet-repeats. The are duplications of "cryptic repeats" composed of mixed synonymous codons and are thought to result from unequal crossing over (Warren, S.T., Science 275, 408-409 (1997)). In the case of OPMD, slippage during replication causing a reiteration of the GCG codon is a more likely mechanism (Wells, D.R., J. Biol. Chem. 271, 2875-2878 (1996)).

Different observations converge to suggest that a gain of function of PAB II may cause the accumulation of nuclear filaments observed in OPMD (Tome, F.M.S. & Fardeau, Acta Neuropath. 49, 85-87 (1980)). PAB II is found mostly in dimeric and oligomeric forms (Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). It is possible that the polyalanine tract plays a role in polymerization. Polyalanine stretches have been found in many other nuclear proteins such as the HOX pro-teins, but their function is still unknown (Davies, S.W. et al., Cell 90, 537-548 (1997)). Alanine is a highly hydrophobic amino acid present in the cores of proteins. In dragline spider silk, polyalanine stretches are thought to form B-sheet structures impor-tant in ensuring the fibers' strength (Simmons, A.H. et al., Science 271, 84-87 (1996)). Polyalanine

oligomers have also been shown to be extremely resistant to chemical denaturation and enzymatic degradation (Forood, B. et al., Bioch. and Biophy. Res. Com. 211, 7-13 (1995)). One can speculate that PAB II oligomers comprised of a sufficient number of mutated molecules might accumulate in the nuclei by forming undegradable polyalanine rich macromolecules. The rate of the accuùmulation would then depend on the ratio of mutated to non-mutated protein. The more severe phenotypes observed in homozygotes for the (GCG)9 mutations and compound heterozygotes for the (GCG)9 mutation and (GCG)7 allele may correspond to the fact that in these cases PAB II oligomers are composed only of mutated proteins. The ensuing faster filament accumulation could cause accelerated cell death. The recent descrip-tion of nuclear filament inclusions in Huntington's disease, raises the possibility that "nuclear toxicity" caused by the accumulation of mutated homopolymeric domains is involved in the molecular pathophysiology of other triplet-repeat diseases (Davies, S.W. et al., Cell 90, 537-548 (1997); Scherzinger, E. et al., Cell 90, 549-558 (1997); DiFiglia, M. et al., Science 277, 1990-1993 (1997)). Future immunocytochemical and expression studies will be able to test this patho-physiological hypothesis and provide some insight into why certain muscle groups are more affected while all tissues express PAB II.

Methods

Contig and cDNA selection

The cosmid contig was constructed by standard cosmid walking techniques using a gridded chromosome 14-specific cosmid library (Evans, G.A. et al., Gene 79, 9-20 (1989)). The cDNA clones were isolated by cDNA selection as previously described (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of transcribed sequences (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

Cloning of the PAB II gene

Three cDNA clones corresponding to PAB II were sequenced (Sequenase, USB). Clones were verified to map to cosmids by South-ern hybridization. The 8 kb HindIII restriction frag-ment was subcloned from cosmid 166G8 into pBluescriptII (SK) (Stratagene). The clone was sequenced using prim-ers derived from the bPABII gene and human EST sequences. Sequencing of the PAB II introns was done by primer walking.

PAB II mutation screening and sequencing

All cases were diagnosed as having OPMD on clinical grounds (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)). RT-PCR- and genomic SSCP analyses were done using stan-dard protocols (Lafrenière, R.G. et al., Nat. Genet. 15, 298-302 (1997)). The primers used to amplify the PAB II mutated region were: 5'-CGCAGTGCCCCGCCTTAGA-3' (SEQ ID NO:4) and 5'-ACAAGATGGCGCCGCCGCCCCGGC-3' (SEQ ID NO:5). PCR reactions were performed in a total volume of 15 µl containing: 40 ng of genomic DNA; 1.5 µg of BSA; 1 µM of each primer; 250 µM dCTP and dTTP; 25 µM dATP; 125 µM of dGTP and 125 µM of 7-deaza-dGTP (Pharmacia); 7.5% DMSO; 3.75 µCi[^{35}S]dATP, 1.5 unit of Taq DNA polymerase and 1.5 mM MgCl2 (Perkin Elmer). For non-radioactive PCR reactions the [^{35}S]dATP was replaced by 225 µM of dATP. The amplification procedure consisted of an initial denaturation step at 95°C for five minutes, followed by 35 cycles of denaturation at 95°C for 15 s, annealing at 70°C for 30 s, elongation at 74°C for 30 s and a final elongation at 74°C for 7 min. Samples were loaded on 5% polyacrylamide denaturing gels. Following electrophoresis, gels were dried and autoradiographs were obtained. Sizes of the inserts were determined by comparing to a standard M13 sequence (Sequenase™, USB). Fragments used for sequencing were gel-purified. Sequencing of the mutated fragment using the Amplicycle kit™ (Perkin Elmer) was

done with the 5'-CGCAGTGCCCCGCCTTAGAGGTG-3' (SEQ ID NO:6) primer at an elongation temperature of 68°C.

Stability of (GCG)-repeat expansions

The meiotic stability of the (GCG)9-repeat was estimated based on a large French Canadian OPMD cohort. It had been previously established that a single ancestral OPMD carrier chro-mosome was introduced in the French Canadian population by three sisters in 1648. Seventy of the seventy one French Canadian OPMD families tested to date segregate a (GCG)9 expansion. However, in family F151, the affected brother and sister, despite sharing the French Canadian ancestral haplotype, carry a (GCG)12 expansion, twice the size of the ancestral (GCG)9 mutation (Fig. 2C). In this founder effect study, it is estimated that 450 (304-594) historical meioses shaped the 123 OPMD cases belonging to 42 of the 71 enrolled families. The screening of the full set of participants allowed an identification of another 148 (GCG)9 carrier chromosomes. Therefore, it is estimated that a single mutation of the (GCG)9 expansion has occurred in 598 (452-742) meioses.

Genotype-phenotype correlations

176 carriers of at least one copy of the (GCG)9 mutation were examined during the early stage of the linkage study. All were asked to swallow 80 cc of ice-cold water as rapidly as possible. Testing was stopped after 60 seconds. The swallowing time (st) was validated as a sensitive test to identify OPMD cases (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995); Bouchard, J.-P. et al., Can. J. Neurol. Sci. 19, 296-297 (1992)). The st values for 76 (GCG)6 homozygotes normal controls is illustrated in Fig. 3. Analyses of variance were computed by two-way ANOVA (SYSTAT package). For the (GCG)9 homozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 35-40 ($P < 10^{-5}$). For the (GCG)9 and (GCG)7 compound

heterozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 45-65 ($P < 10^{-5}$).

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any vari-ations, uses, or adaptations of the invention follow-ing, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.